

10/535003

JC14 Rec'd PCT/PTO 12 MAY 2005

Express Mail Label No. EV582716495US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: JEONG-BUM PYUN , ET AL)
)
)
For: DATA SEARCHING METHOD AND INFORMATION DATA)
SCRAPPING METHOD USING INTERNET)

CLAIM FOR PRIORITY

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450


Dear Sir:

Applicant hereby claims the benefits of the filing date of November 12, 2002 to Korean Patent Application No. 10-2002-0070187 under provisions of 35 U.S.C. 119 and the International Convention for the protection of Industrial Property.

If any fees are due with regard to this claim for priority, please charge them to Deposit Account No. 06-1130 maintained by Applicant's attorneys.

Respectfully submitted,

CANTOR COLBURN LLP

By 
Soonja Bae
Registration No. (See Attached)
Cantor Colburn LLP
55 Griffin Road South
Bloomfield, CT 06002
PTO Customer No. 23413
Telephone: (860) 286-2929
Facsimile: (860) 286-0115

Date: May 12, 2005

BEST AVAILABLE COPY

Rec'd PCT/PTC 12 MAY 2005
PCT/KR 03/02323
RO/KR 31.10.2003 #2

REC'D 25 NOV 2003
WIPO PCT



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2002-0070187
Application Number

출원 년 월 일 : 2002년 11월 12일
Date of Application NOV 12, 2002

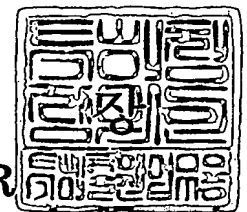
출원인 : 편정범 외 1명
Applicant(s) PYUN, JEONG BUM, et al.



2003 년 10 월 31 일

특 허 청

COMMISSIONER



PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

【서지사항】

【서류명】	특허출원서		
【권리구분】	특허		
【수신처】	특허청장		
【제출일자】	2002.11.12		
【발명의 명칭】	인터넷을 이용한 정보 자료 스크랩 방법		
【발명의 영문명칭】	Information Materials Scrapping Method using Internet		
【출원인】			
【성명】	편정범		
【출원인코드】	4-2002-042224-8		
【출원인】			
【성명】	박원준		
【출원인코드】	4-2002-021590-1		
【대리인】			
【명칭】	청운특허법인		
【대리인코드】	9-2002-100001-8		
【지정된변리사】	이철 , 이인실 , 염승운 , 최재승		
【포괄위임등록번호】	2002-083090-9		
【포괄위임등록번호】	2002-083082-5		
【발명자】			
【성명】	편정범		
【출원인코드】	4-2002-042224-8		
【발명자】			
【성명】	박원준		
【출원인코드】	4-2002-021590-1		
【심사청구】	청구		
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 청운특허법인 (인)		
【수수료】			
【기본출원료】	20	면	29,000 원
【가산출원료】	8	면	8,000 원
【우선권주장료】	0	건	0 원
【심사청구료】	9	항	397,000 원

0070187

출력 일자: 2003/11/8

【합계】	434,000 원
【감면사유】	개인 (70%감면)
【감면후 수수료】	130,200 원
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 인터넷을 이용한 정보 자료 스크랩 방법에 관한 것으로, 사용자용 컴퓨터에 접속된 온라인을 통하여 검색사이트의 검색 기능을 이용하여 원하는 검색 정보의 키워드를 입력하여 검색하는 단계; 상기 사용자용 컴퓨터에 설정된 프로그램이 자동으로 HTTP 프로토콜을 통하여 검색 사이트의 웹서버에 접속하는 단계; 상기 접속된 검색 사이트의 웹서버에 검색을 위한 쿼리(Query)를 송신하는 단계; 상기 웹서버가 수신한 쿼리에 대한 결과로서 연계된 하나 이상의 데이터서버로부터 검색한 결과물을 인터넷을 통하여 사용자용 컴퓨터로 송신하는 단계; 상기 검색된 자료를 HTTP 프로토콜을 이용하여 다운로드받는 단계; 상기 다운로드한 정보 중에서 불필요한 정보를 제거하는 단계; 상기 불필요한 정보가 제거된 결과를 로컬 저장매체에 저장하는 단계; 상기 로컬 저장매체에 정보를 사용자용 컴퓨터에 포함된 프로그램에 의하여 편집, 가공 및 관리하는 단계를 포함하여 로컬의 저장매체에 사용자가 수집, 가공 및 저장함으로써 정보 검색시간 및 자료관리를 용이하게 한 것이다.

【대표도】

도 2

【색인어】

인터넷, 스크랩, 신문, 서버, 키워드, 웹사이트

【명세서】**【발명의 명칭】**

인터넷을 이용한 정보 자료 스크랩 방법{Information Materials Scrapping Method using Internet}

【도면의 간단한 설명】

도 1은 일반적인 인터넷 사용자의 정보 수집 절차를 나타낸 흐름도,

도 2는 본 발명에 따른 인터넷을 이용한 정보 자료 스크랩 시스템의 구성도,

도 3은 본 발명에 따른 인터넷을 이용한 정보 자료 검색 및 수집절차를 나타낸 흐름도,

도 4는 본 발명에 따른 인터넷을 이용한 정보자료의 관리절차를 나타낸 흐름도,

도 5는 본 발명에 따른 프로그램의 메인 화면을 나타낸 도면,

도 6은 본 발명에 의한 프로그램을 통하여 검색할 키워드를 입력하는 과정을 나타낸 도면,

도 7은 본 발명에 의한 프로그램을 이용하여 실제 자료를 검색하는 과정을 나타낸 도면,

도 8은 본 발명에 의한 프로그램을 이용하여 등록된 키워드 리스트를 나타낸 도면,

도 9는 본 발명에 의한 프로그램을 이용하여 실제로 검색된 자료의 리스트를 나타낸 도면,

도 10은 본 발명에 의한 프로그램을 이용하여 검색한 내용을 확인하는 모습을 나타낸 도면,

도 11은 본 발명에 의한 프로그램을 이용하여 검색한 내용을 하나의 파일로 변환한 상태를 나타낸 도면.

♣ 도면의 주요부분에 대한 부호의 설명 ♣

10: 사용자용 컴퓨터 12: 데이터 프로세싱 엔진 소프트웨어

20: 모니터 30: 입력수단

40: 저장매체 50: 인터넷

60: 웹서버 70: 데이터서버

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

17> 본 발명은 인터넷을 이용한 정보 자료 스크랩 방법 및 그 시스템에 관한 것으로, 더욱 상세하게는 인터넷을 통하여 제공되는 신문, 잡지 또는 웹 문서 등과 같이 자료 보관의 의미를 가지는 유용한 데이터를 로컬의 저장매체에 사용자가 수집, 가공 및 저장함으로써 정보 검색시간 및 자료관리를 용이하게 하는 인터넷을 이용한 정보 자료의 스크랩 방법 및 그 시스템에 관한 것이다.

18> 종래에 온라인을 통한 각종 자료의 검색 및 수집은 도 1의 흐름도를 참조하면 다음과 같다.

19> 먼저, 단계(S1)에서 사용자가 검색을 원하는 사이트(예를 들어, 신문이나 잡

지 사이트 등)에 접속을 한다. 이때 접속은 온라인으로 접속된 인터넷을 통하여 검색을 원하는 사이트에 접속을 하는 것이다. 단계(S2)에서는 원하는 사이트로의 접속이 이루어지면 사용자가 검색하고자 하는 키워드를 입력한다. 즉 해당하는 사이트의 검색 키워드 입력란에 검색을 원하는 키워드를 입력하는 것이다. 단계(S3)에서는 단계(S2)에서 입력한 키워드에 대한 검색이 수행되면 검색된 정보에 관한 리스트가 PC의 화면상에 출력된다.

<20> 단계(S4)에서 사용자는 PC의 화면상에 출력된 리스트를 참조하여 원하는 하나의 링크를 클릭하여 확인한다. 이때, 사용자는 리스트 중에서 임의의 링크를 클릭하거나 사용자가 원하는 가장 유력한 링크를 클릭하여 볼 수 있을 것이다. 단계(S5)에서는 상기 단계(S4)에서 클릭하여 열어본 링크의 내용을 읽어보고 사용자가 원하는 필요한 내용을 포함하고 있는 지를 판단한다. 단계(S6)에서는 사용자가 필요로 하는 정보가 포함되어 있을 경우에 입력장치인 키보드 또는 마우스를 이용하여 해당하는 내용을 선택한 후에 복사를 한다. 이렇게 복사한 내용은 단계(S7)에서 사용자가 원하는 워드프로세서(예를 들어, 한글 또는 MS워드 등)에 문자로서 붙여 넣는다.

<21> 이러한 과정, 즉 단계(S4)에서 단계(S7)까지의 작업을 순차적으로 반복하여 수행하게 된다. 이를 통하여 사용자가 원하는 정보를 수집하고 수집된 내용을 편집할 수 있다. 그리고 단계(S8)에서 사용자의 선택에 의하여 더 이상의 확인할 내용이 없는 지를 판단한 후에 단계(S9)에서 다른 검색사이트를 이용하여 동일한 작업을 할 것인지를 판단한다. 그리고 다른 검색사이트를 통한 정보 수집을 원하지 않을 경우에는 정보 수집 작업을 종료하게 된다.

<22> 이와 같이 종래에 상기와 같은 과정을 통해서 얻어진 자료는 사용자의 손에 익숙한 워드 프로세서 등으로 이미지파일이나 텍스트파일 등으로 저장되고 관리되는 경우가 대부분이다.

<23> 그러나 실제로 이러한 작업을 할 경우에 몇 가지 문제점이 야기된다. 그 중에서도 가장 큰 문제점은 이러한 작업에 소요되는 시간이 상당히 크다는 것이다. 실제로 온라인 검색에 소요되는 시간을 가상적으로 계산해 보기로 하면, 단 여기서는 현재 가장 많이 사용되는 초고속 인터넷(ADSL) 이상의 환경을 대상으로 하였을 경우에 검색 사이트의 접속하는 시간 약 5~10초, 키워드의 입력시간 약 5~10초, 검색 결과를 기다리는 시간(각종 광고나 관련 링크 또는 선택창 등의 부수적인 자료 포함) 약 2~20초, 검색 결과 중에서 사용자가 하나의 항목을 선택하여 클릭하는 시간 약 3~5초, 내용을 확인하고 필요여부를 판단하는 시간 약 10~20초, 필요할 경우에 내용을 선택해서 복사하는 시간 약 10초 내외, 그리고, 선택한 내용을 워드 프로세서 등에 붙이기를 하는 시간 약 5초가 소요된다.

<24> 이러한 일련의 과정을 거치게 되는데, 그 과정을 분석해 보면, 우선 정보를 수집하는 시간이 상당히 많이 소요된다는 것을 알 수 있다. 그 첫 번째 이유로는 사람과 네트워크, 그리고 PC 등, 여러 작업 주체의 기능이 혼재됨으로 작업 주체간 전환되는 시간이 많이 소요된다는 점이다. 즉 사용자의 조작 → 네트워크를 통한 접속의 대기 → 사용자의 조작 → PC 기능 → 사용자의 판단 → 사용자의 조작 등으로 이루어지기 때문이다.

<25> 또한, 시간이 많이 소요되는 두 번째 이유로는 유용한 콘텐츠가 포함된 웹 화면에는 불필요한 광고, 링크 또는 이미지 등이 일반적으로 40~50개 가량 포함되어 있어 본문의 내용 확인을 위해서는 이러한 불필요한 내용이 포함된 화면이 뜨는 시간을 기다릴 수밖에 없다는 점이다. 또한 일반적으로 하나의 사이트만을 대상으로 검색을 수행하지 않는 관계로 다른 사이트에 접속해서 이러한 과정을 반복적으로 수행하여야 한다.

<26> 또한, 시간이 많이 소요된다는 것 이외에도 정보의 수집과정이 반복적이고 지루하다는 단점이 있다.

- <27> 또한, 이러한 반복적 과정에서 정보의 누락이나 중복 등이 발생할 수 있으며, 이 경우에 재검색 등의 불필요한 작업이 추가로 발생될 수 있다. 또한 만일 이러한 작업이 매일 또는 자주 수행하는 경우에는 불편사항은 더욱 늘어나게 될 것이다.
- <28> 현재 이러한 불편사항을 어느 정도는 해결한 메타엔진의 소프트웨어들이 등장하고 있으나, 이러한 소프트웨어들 역시 검색 결과를 한 곳에 모아 놓은 수준, 즉 검색 결과가 있는 URL(Uniform Resource Locator; URL은 인터넷에서 접근 가능한 자원의 주소를 일관되게 표현할 수 있는 형식을 말한다.)만을 표시하는 정도의 서비스만을 제공하고 있는 실정이다.
- <29> 더욱이 국내특허공개 제2001-10807호(인터넷을 이용한 뉴스정보 스크랩의 방법 및 시스템)는 신문기사 스크랩과 거의 동일한 형태로 인터넷을 이용하여 정보출처가 기록된 신문의 뉴스기사, 광고, 광고 등과 같은 관심이 있는 뉴스정보를 이미지파일과 텍스트파일로 제공하는 것으로, 검색된 결과는 사용자만의 데이터베이스 저장공간을 제공하는 것이다. 따라서, 사용자가 스크랩한 정보를 다시 볼 경우에는 인터넷을 접속하여 검색된 결과가 저장된 데이터베이스의 저장공간을 열람하여야 하는 것으로는 이는 자신의 고유한 서버를 필요로 하는 것이다.
- <30> 또한, 국내특허공개 제2001-102786호 및 국내특허공개 제2002-26082호의 경우에도 스크랩서버나 데이터베이스 등의 저장공간을 통하여 해당하는 정보를 분류, 수정, 검색하는 서비스를 제공하는 것을 목적으로 하고 있는 것인바, 이 기술은 모두 인터넷을 통하여 해당하는 서버 및 데이터베이스에 가공된 정보를 열람할 수 있도록 되어 있어 오프라인 상태에서는 수집한 정보를 열람할 수 없도록 되어 있는 단점이 있다.

【발명이 이루고자 하는 기술적 과제】

<31> 이상의 문제점을 해결하기 위해서는 기술의 복합적인 제공을 필요로 하고, 이는 다음의 몇 가지 기능들을 필요로 한다.

<32> 첫째, 일괄 검색 기능으로서, 여러 개의 검색 사이트를 한 번에 검색해서 그 결과를 한꺼번에 볼 수 있는 기능이 필요하다. 둘째, 검색 결과의 적절한 가공으로서, 현재 온라인 정보의 경우 각종 배너 및 광고로 인해 그 내용을 확인하는데 보다 많은 시간이 소요될 뿐만 아니라 자료의 보관 및 관리 등이 어려워진다. 이러한 불필요한 자료를 적절히 처리해 주는 기능이 필요하다. 셋째, 데이터의 확인 시의 속도 향상으로서, 특히 검색된 내용이 많은 경우 보다 빨리 그 내용을 확인할 수 있어야 한다. 즉 몇 천건 이상의 데이터를 확인할 경우에 현재의 온라인 상에서의 확인처럼 내용을 확인할 때마다 몇 초 이상이 소요된다면 그 시간의 소모가 커질 수밖에 없을 것이다. 이러한 내용의 확인을 보다 빠르게 할 필요가 있다. 넷째, 데이터 관리의 용이성으로서, 확인한 내용은 그 필요여부에 따라 보다 편리하게 관리될 수 있어야 한다. 즉 보관이 필요한 내용은 보관을, 필요치 않은 내용은 쉽게 삭제할 수 있어야 한다. 또한 워드 프로세서로의 변환 등 관리의 용이성이 필요하다. 그리고 다섯째, 검색된 내용의 자동적인 갱신 기능으로서, 검색된 내용을 사용자가 원할 경우에 일정한 주기로 자동으로 갱신할 수 있는 기능이 필요하다. 현대의 정보는 새로운 정보가 아니면 그 가치가 떨어질 수밖에 없으므로 항상 새로운 자료로 그 내용을 유지해야 하는데 이러한 사항을 모두 자동으로 처리할 경우 사용자의 시간적, 육체적, 정신적 만족도는 더욱 커질 수 있을 것이다.

<33> 본 발명은 상기와 같은 문제점을 해소하기 위한 것으로, 인터넷을 통한 신문, 잡지 또는 웹 문서 등과 같이 정보로서 가치가 있는 콘텐츠에 포함된 불필요한 광고나 이미지 또는 관련

링크사이트 등을 제거하고 꼭 필요한 텍스트와 이미지만을 획득하여 로컬의 저장매체에 저장할 수 있는 형태로 변환 및 저장한 후에 이를 사용자 편의에 맞도록 재수정할 수 있도록 함과 동시에 하나 이상의 파일로 변환 및 저장할 수 있도록 하는 인터넷을 이용한 정보 자료의 스크랩 방법과 그 시스템을 제공하기 위한 것이 목적이다.

【발명의 구성 및 작용】

34> 본 발명은 상기의 목적을 달성하기 위하여, 사용자용 컴퓨터에 접속된 온라인을 통하여 검색사이트의 검색 기능을 이용하여 원하는 검색 정보의 키워드를 입력하여 검색하는 단계; 상기 사용자용 컴퓨터에 설정된 프로그램이 자동으로 HTTP 프로토콜을 통하여 검색 사이트의 웹 서버에 접속하는 단계; 상기 접속된 검색 사이트의 웹서버에 검색을 위한 쿼리(Query)를 송신하는 단계; 상기 웹서버가 수신한 쿼리에 대한 결과로서 연계된 하나 이상의 데이터서버로부터 검색한 결과물을 인터넷을 통하여 사용자용 컴퓨터로 송신하는 단계; 상기 검색된 자료를 HTTP 프로토콜을 이용하여 다운로드받는 단계; 상기 다운로드한 정보 중에서 불필요한 정보를 제거하는 단계; 상기 불필요한 정보가 제거된 결과를 로컬 저장매체에 저장하는 단계; 상기 로컬 저장매체에 정보를 사용자용 컴퓨터에 포함된 프로그램에 의하여 편집, 가공 및 관리하는 단계를 포함하여 이루어진 인터넷을 이용한 정보 자료 스크랩 방법을 제공한다.

35> 이하 첨부된 도면을 참조하여 본 발명에 관하여 상세하게 설명하면 다음과 같다.

36> 도 2는 본 발명에 따른 인터넷을 이용한 정보 자료 스크랩 시스템의 구성도로, 인터넷에 연결된 로컬의 사용자용 컴퓨터에 설치된 데이터 프로세싱 엔진 소프트웨어(Data Processing Engine Software)가 인터넷을 통해서 검색하고자 하는 웹서버(Web Server)에 접속하여 검색 결과를 얻고, 얻어진 검색 결과를 로컬 저장매체(플로피디스크, 하드디스크, 콤팩트디스크 또는 플래쉬 메모리 등)에 저장하는 시스템의 구성도를 나타낸 것이다.

- <37> 본 발명의 사용자용 컴퓨터(10)는 데스크탑 컴퓨터이거나 휴대용 컴퓨터 등이고, 온라인을 통한 통신이 가능한 것이어야 한다. 그리고 사용자용 컴퓨터(10)에는 데이터 프로세싱 엔진 소프트웨어(12)가 설치되어 있어야 한다. 데이터 프로세싱 엔진 소프트웨어(12)는 인터넷을 통하여 검색을 하고 검색된 자료를 저장매체에 저장시키는 기능을 포함한 엔진으로 프리웨어 또는 셰어웨어일 수 있다. 또한, 데이터 프로세싱 엔진 소프트웨어는 다운로드되어 로컬 저장매체에 저장된 각각의 파일의 집합을 하나 이상의 파일로 변환 및 저장해 주는 기능을 포함하고 있다.
- <38> 모니터(20)는 검색된 결과를 가시적으로 표시하거나 입·출력장치의 입·출력상태를 표시하는 것이다. 입력수단(30)은 검색에 필요한 키워드를 입력하거나 검색결과에 대한 편집 등을 할 수 있는 것으로 키보드나 마우스 등이 포함된다.
- <39> 저장매체(40)는 상기 데이터 프로세싱 엔진 소프트웨어(12)로부터 검색된 데이터를 저장하기 위한 것으로, 플로피디스크, 하드디스크, 콤팩트디스크 또는 플래쉬 메모리 등이 포함된다.
- <40> 웹서버(60)는 로컬의 사용자용 컴퓨터(10)로부터 인터넷(50)을 통하여 해당하는 웹사이트(신문사 또는 잡지사 또는 기타 각종의 정보를 제공하는 사이트 등)의 서버이다. 웹서버(60)에는 데이터, 이미지 데이터 또는 각종의 데이터 정보를 제공하는 복수의 데이터서버(70, 72, 74)가 연계되어 있다.
- <41> 본 발명의 기본적인 배경 기술은 스크린 스크래핑(Screen Scrapping)이다. 여기서 스크린 스크래핑은 인터넷 웹사이트의 내용을 읽어와서 그 중에서 필요한 내용만을 발췌해 내는 기술이다.

- <42> 즉 스크린 스크래핑의 예를 들면, 날씨 제공 사이트에서 날씨 정보를 읽어와서 사용하거나 뉴스 제공 사이트에서 뉴스를 읽어와서 사용하거나 또는 증권 정보 사이트에서 증권 정보를 읽어와서 사용하는 등의 예를 들 수 있을 것이다.
- <43> 본 발명은 상기 스크린 스크래핑 기능을 바탕으로 이루어지는 자료의 검색 및 수집절차는 도 3을 참조하여 설명한다.
- <44> 먼저, 단계(S10)에서 사용자용 컴퓨터(10)에 접속된 온라인을 통하여 검색사이트(예를 들면, 뉴스정보를 제공하는 신문사, 일간지나 월간지 등을 제공하는 잡지사 또는 각종 정보를 제공하는 웹사이트 등)의 검색 기능을 이용하여 원하는 검색 정보의 키워드를 입력하여 검색을 수행한다. 이때 검색하고자 하는 키워드는 입력수단(30)을 이용하여 검색란에 입력한다. 즉 예를 들어, 온라인을 통하여 뉴스정보를 제공하는 신문사 등의 검색 기능을 이용하여 원하는 내용을 검색하는 것이다. 이때 여러 개의 사이트를 동일한 검색어를 이용해서 한번에 검색할 수 있는 통합 검색기능도 제공한다.
- <45> 단계(S11)에서는 상기 사용자용 컴퓨터(10)에 설정된 프로그램이 자동으로 HTTP 프로토콜을 통하여 인터넷(50)을 거쳐 검색 사이트의 웹서버(60)에 접속한다.
- <46> 상기 HTTP(Hypertext Transfer Protocol; 하이퍼텍스트 전송 규약)은 웹 상에서 파일(텍스트, 그래픽 이미지, 사운드, 비디오 그리고 기타 멀티미디어 파일)을 주고받는데 필요한 프로토콜로서 TCP/IP(Transmission Control Protocol/Internet Protocol; TCP/IP는 인터넷의 기본적인 통신 프로토콜)와 관련된 하나의 응용 프로토콜이다.
- <47> 그리고 단계(S12)에서 상기 접속된 검색 사이트의 웹서버(60)에 검색을 위한 쿼리(Query)를 송신하고, 단계(S13)에서 상기 웹서버(60)가 수신한 쿼리에 대한 결과로서 연계된

하나 이상의 데이터서버(70, 72, 74)로부터 검색한 결과물을 인터넷(50)을 통하여 사용자용 컴퓨터(10)로 송신한다.

<48> 상기 검색된 내용을 이용하여 그 실제 내용을 읽어들인다. 즉 검색된 결과는 그 실제내용을 연결하기 위한 하이퍼링크(Hyperlink)인 경우가 대부분이기 때문이다. 따라서 본 발명은 검색된 링크 정보를 이용하여 실제 정보 내용을 읽어들이는 작업까지 수행하게 된다. 이 과정에서 스크린 스크래핑 기술이 사용되게 된다. 즉 실제 내용과 연결된 링크를 분석해서 찾아내야 하는데 이 부분에서 스크린 스크래핑 기술이 사용되게 된다. 즉 단계(S14)에서 상기 검색된 자료를 HTTP 프로토콜을 이용하여 다운로드받는다.

<49> 그리고 단계(S15)에서 상기 다운로드한 정보 중에서 불필요한 정보를 제거하게 된다. 이때 읽어들인 정보를 적절한 형태로 변형시키게 되는데, 여기서 적절한 형태는 다음과 같은 과정을 거친 다음의 형태를 말한다.

<50> 불필요한 정보의 제거로서 각종 광고 정보 및 불필요한 링크 등은 제거하고, 이미지 데이터 링크 등의 변환으로서 내용에 필요한 이미지인 경우에 그 링크 내역을 온라인 링크에서 오프라인 링크로 변경한다. 이때 링크의 변환 방법은 다음과 같다.

<51> 실제 이미지의 이름을 추출한다. 예를 들어, <http://www.test.com/test.jpg> 파일의 경우에는 "test.jpg" 라는 이름을 추출해 내게 된다. 추출된 이미지 이름의 앞 부분에 상대 위치를 추가한다. 이때 상대 위치를 img 폴더가 된다. 즉 test.jpg 파일이 있는 경우 img/test.jpg가 되게 된다. 그리고 절대 링크에 있는 이미지 파일을 img 폴더에 다운로드 한다. 이로써 이미지를 포함한 로컬데이터의 생성이 가능하게 된다.

- <52> 또한, 필요 정보의 추가로서 각종 HTML 링크를 추가하게 된다. 불필요한 정보를 제거하는 과정에서 주로 앞이나 뒤의 정보는 제거되고 중간 부분의 정보만이 남게 되는 경우가 많고, 이 과정에서 반드시 필요한 태그가 삭제되는 경우가 있다. 즉 HTML 문서임을 나타내는 <html> 태그 등이 제거될 수 있는데 이러한 중요 태그 정보를 추가하게 된다.
- <53> 단계(S16)에서는 상기 불필요한 정보가 제거된 결과를 로컬 저장매체(40)에 저장한다. 즉 변형된 정보를 로컬의 저장매체(40)에 저장하고, 실제 내용은 개별 파일을 만들어서 저장한다. 그리고 그 링크 내역을 데이터베이스에 저장한다. 이렇게 내용과 링크를 분리함으로써 데이터 검색 속도를 향상시킨다. 또한 데이터베이스에 문제가 생겼을 경우에 그 피해를 최소화한다. 또한 저장된 개별 파일을 따로 사용할 수도 있게 된다.
- <54> 단계(S17)는 상기 로컬 저장매체(40)에 정보를 사용자용 컴퓨터(10)에 포함된 프로그램(12)에 의하여 편집, 가공 및 관리한다.
- <55> 상기 단계(S17)는 도 4의 로컬 저장매체(40)에 저장된 자료를 관리하는 과정을 나타낸 흐름도이다. 즉 단계(S20)에서는 로컬 저장매체(40)에 저장된 자료를 읽어낸다. 그리고 단계(S21)에서 읽어들이는 자료의 내용을 확인하고, 단계(S22)에서 자료의 내용이 필요한 내용인지를 판단한다. 필요하지 않은 내용일 경우에는 단계(S23) 및 (S24)과 같이 입력수단(30)을 통하여 삭제키를 입력하여 확인한 내용을 삭제한다. 하지만 필요한 내용일 경우에는 단계(S25)에서 확인하지 않은 자료가 있는지를 판단한다. 따라서, 단계(S22) 내지 단계(S25)의 과정을 반복하여 내용 확인을 하게 된다.
- <56> 단계(S18)에서는 등록된 다른 검색사이트를 더 검색할 것인지를 판단하여 단계(S11) 내지 단계(S17)를 반복적으로 수행한다.

- <57> 상기 단계(S17)와 단계(S18)는 사용자의 요구에 의하여 그 순서가 바뀔 수 있다. 즉 저장매체에 저장된 데이터를 가공한 후에 등록된 다른 검색사이트를 검색할 수도 있고, 다른 검색사이트를 검색한 후에 저장매체에 저장된 데이터를 가공할 수도 있다.
- <58> 상기 과정을 거쳐 저장된 정보는 사용자가 용이하게 삭제하거나 합치기 등의 기능을 통해 관리될 수 있도록 하고, 저장된 정보는 백업 기능을 통해 로컬의 저장매체에서 다른 저장매체로 쉽게 저장 및 복원이 가능하게 한다. 그리고 자동 갱신 기능을 통해 지정된 검색어에 관련된 정보를 일정 주기마다 자동으로 갱신할 수 있도록 하여 사용자의 편의를 도모한다.
- <59> 도 5는 본 발명을 이용한 프로그램의 메인 화면을 나타낸 것으로, 화면의 왼쪽에는 사용자가 검색한 키워드가 나열되어 있고, 현재 바둑기보를 선택하였을 경우에 검색된 결과가 상단에 표시되고, 그 하단에는 현재 결과에 대하여 저장된 기사제목 및 관련정보가 표시되게 된다.
- <60> 도 6은 본 발명에 의한 프로그램을 통하여 검색할 키워드를 등록하는 과정을 나타낸 것으로, 검색할 키워드, 검색대상, 검색기간 지정 등을 입력하는 것이다. 한번 등록된 키워드는 사용자의 선택에 따라 삭제될 수 있다.
- <61> 도 7은 본 발명에 의한 프로그램을 이용하여 실제 자료를 검색하는 화면을 나타낸 것으로, 전체 검색상황, 현재 사이트 검색상황, 현재 사이트 저장상황, 현재 사이트 및 검색된 자료의 수 등이 가시적으로 표시되면서 검색이 이루어진다.
- <62> 도 8은 본 발명에 의한 프로그램을 이용하여 등록된 키워드 리스트를 나타낸 것으로, 사용자가 등록한 검색을 위한 키워드가 등록된 상태를 나타낸 것이고, 도 9는 본 발명에 의한 프로그램을 이용하여 실제로 검색된 자료의 리스트를 나타낸 것으로, 제목, 신문사, 날짜 등이 나타나 있다.

- <63> 도 10은 본 발명에 의한 프로그램을 이용하여 검색한 내용을 확인하는 모습을 나타낸 것으로, 텍스트 및 이미지가 표시된 것이고, 도 11은 본 발명에 의한 프로그램을 이용하여 검색한 내용을 하나의 파일로 변환한 상태를 나타낸 것으로, 마이크로소프트 워드 파일로 변환한 상태를 나타낸 것이다.
- <64> 다음은 본 발명의 실시예로서, 온라인을 통한 뉴스정보를 제공하는 신문, 예를 들어, 조선일보 웹사이트를 대상으로 본 발명에 의한 정보 검색 프로그램을 활용한 것이다.
- <65> 그 결과 실제 검색에 소요된 시간은 일반적인 방법, 즉 웹사이트에 접속해서 검색한 후에 내용을 확인하는 방법보다 대략 500%이상의 효율을 나타냈다. 특히 검색 결과가 많은 경우에는 그 효율이 더욱 높아졌다.
- <66> 자료 검색 효율의 계산 예로서, 사용자용 컴퓨터에 Windows2000 운영체제의 초고속 인터넷환경(xDSL)에서 수행하였을 경우이다.
- <67> 만약, "창업"이라는 키워드를 이용해서 검색을 수행하였을 경우에 약 6,000 건이 넘는 데이터가 검색되는데 이 사항을 일반적인 방법으로 모두 확인하려면, 1개당 확인 후 판단 시간이 약 5초가 걸린다면 $5\text{초} \times 6,000\text{건} = \text{약 } 8.3\text{시간}$ 이 소요된다.
- <68> 그리고 이 중에서 필요한 자료를 복사하고 저장하는 시간은 이보다 최소 3~4배 이상 걸릴 것이다. 따라서 최소 20시간 이상의 시간이 소요됨을 알 수 있다.
- <69> 그러나 본 발명의 데이터 프로세싱 엔진 소프트웨어가 로컬의 사용자용 컴퓨터에 설치된 경우에는 6,000건을 검색하는데 걸린 시간이 약 20~30분(초고속 인터넷의 상태에 따라 틀림)이고, 확인하는데 걸리는 시간은 개당 $1.5\text{초} \times 6,000 = \text{약 } 2\text{시간 } 30\text{분}$ 이 소요된다. 또한 확인과

삭제, 그리고 저장이 동시에 이루어지므로 자료의 복사 및 저장 시간이 필요치 않게 된다. 따라서 전체 소요 시간은 약 3시간 가량이 되게 된다.

<70> 즉 객관적인 비교로 종래의 방식으로는 약 20시간이 걸리던 것이 약 3시간이면 충분하므로 그 효율로는 약 600%가 넘는 시간적 효율성을 보이게 된다.

<71> 또한 본 발명의 경우, 검색이 이루어지는 시간 동안은 다른 작업을 할 수 있으므로 실제 사용자가 소요하는 시간은 그보다 더 적다고 할 수 있겠다.

【발명의 효과】

<72> 상술한 바와 같이 본 발명의 인터넷을 이용한 정보 자료 스크랩 방법은 분야별, 이용 대상별로 다양한 활용도를 가질 수 있고, 일반 기업의 기획부서나 홍보 부서의 경우 해당 제품, 경쟁 제품, 시장 동향 등의 자료를 조사, 보관하는데 아주 효율적으로 이용될 수 있다. 그리고 영업부서의 경우 영업 대상이 되는 회사의 정보 조회 및 산업 동향, 각종 인물의 대한 정보를 조사, 보관하는데 유용하게 사용되어 질 수 있고, 개인 사용자의 경우에는 일반적으로 창업 준비자의 경우 창업에 관련된 정보를 조사하는데 아주 유용하게 이용되어 질 수 있으며, 또한 주식 투자를 할 경우 주식을 소유한 해당 기업의 뉴스 및 동향, 산업의 전반적인 동향에 관련된 정보를 얻는데 사용될 수 있다.

<73> 그리고 학생의 경우에는 각종 리포트 자료 수집시 활용도를 향상시킬 수 있고, 좋아하는 연예인의 기사나 사진 등을 수집하는데 용이하게 활용할 수 있고, 취미생활이나 건강에 관련된 자료를 수집하는데도 사용될 수 있는 등, 검색한 자료를 유용하게 사용할 수 있을 것이다.

<74> 더욱이 본 발명은 데이터 프로세싱 엔진 소프트웨어 등에서 검색된 웹 문서를 최소한의 형태로 줄인 후에 그 모든 내용을 로컬 저장매체에 저장함으로서, 온라인을 통한 인터넷의 연결 유

무와는 상관없이 확인이 가능하고, 검색 및 검색내용의 확인에 소요되는 시간을 최소화하여 검색에 필요한 시간을 절감할 수 있는 효과가 있다. 또한 이미 최소화되어 저장되어 있는 자료들이기 때문에 자료의 삭제 및 합치기 등으로 자료의 관리가 용이한 장점도 있다.

【특허청구범위】**【청구항 1】**

사용자용 컴퓨터에 접속된 온라인을 통하여 검색 사이트의 검색 기능을 이용하여 원하는 검색 정보의 키워드를 입력하여 검색하는 단계;

상기 사용자용 컴퓨터에 설정된 프로그램이 자동으로 HTTP 프로토콜을 통하여 검색 사이트의 웹서버에 접속하는 단계;

상기 접속된 검색 사이트의 웹서버에 검색을 위한 쿼리(Query)를 송신하는 단계;

상기 웹서버가 수신한 쿼리에 대한 결과로서 연계된 하나 이상의 데이터서버로부터 검색한 결과물을 인터넷을 통하여 사용자용 컴퓨터로 송신하는 단계;

상기 검색된 자료를 HTTP 프로토콜을 이용하여 다운로드받는 단계;

상기 다운로드한 정보 중에서 불필요한 정보를 제거하는 단계;

상기 불필요한 정보가 제거된 결과를 로컬 저장매체에 저장하는 단계;

상기 로컬 저장매체에 정보를 사용자용 컴퓨터에 포함된 프로그램에 의하여 편집, 가공 및 관리하는 단계를 포함하여 이루어진 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 2】

제1항에 있어서, 상기 사용자용 컴퓨터에 포함된 프로그램(데이터 프로세싱 엔진 소프트웨어)은 관리하고자 하는 자료의 갱신을 사용자가 지정한 검색어에 관련된 정보를 일정 주기마다 자동으로 갱신이 이루어지도록 하는 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 3】

제1항에 있어서, 상기 불필요한 정보의 제거는 각종 광고 정보 및 불필요한 관련 링크를 제거하는 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 4】

제1항에 있어서, 상기 이미지 데이터 링크 등의 변환은 내용에 필요한 이미지의 경우에 그 링크 내역을 온라인 링크에서 오프라인 링크로 변경하는 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 5】

제1항에 있어서, 상기 검색 정보는 온라인 신문, 온라인 잡지, 웹 문서 중에서 어느 하나인 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 6】

제1항에 있어서, 상기 다운로드받은 자료 중에서 필요하지 않은 태그 부분을 제거하고 필요한 부분만을 저장함으로서 저장 시간 및 저장 공간을 최소화시킨 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 7】

제1항에 있어서, 상기 사용자용 컴퓨터에 포함된 프로그램(데이터 프로세싱 엔진 소프트웨어)은 이미지와 같은 부가 자료를 로컬 저장매체에서 사용하기 위하여 다운로드되어 저장된 HTML 문서의 내용을 자동으로 변경시키는 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 8】

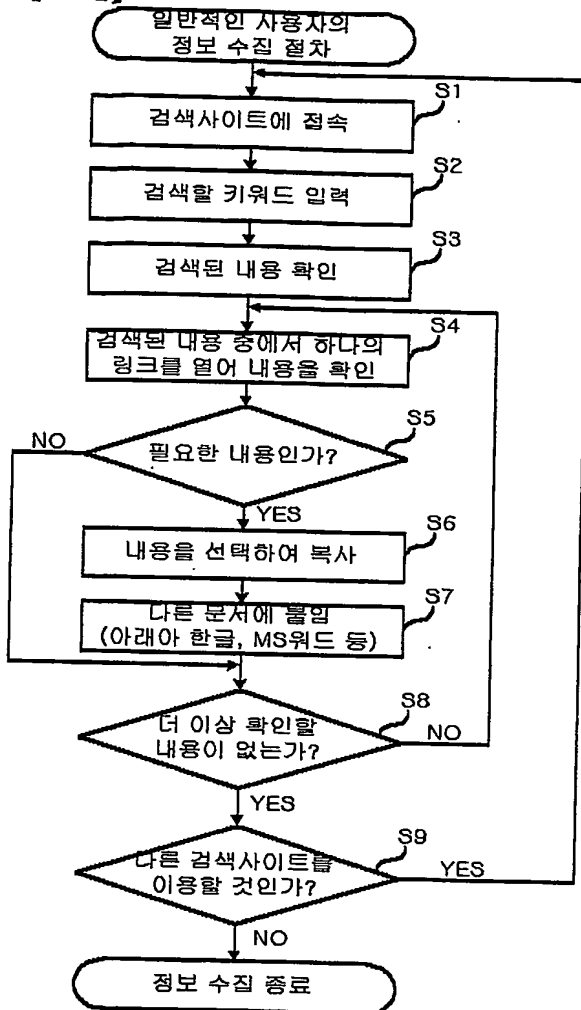
제1항에 있어서, 상기 사용자용 컴퓨터에 포함된 프로그램(데이터 프로세싱 엔진 소프트웨어)은 다운로드되어 로컬 저장매체에 저장된 각각의 파일의 집합을 하나 이상의 파일로 변환 및 저장해 주는 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【청구항 9】

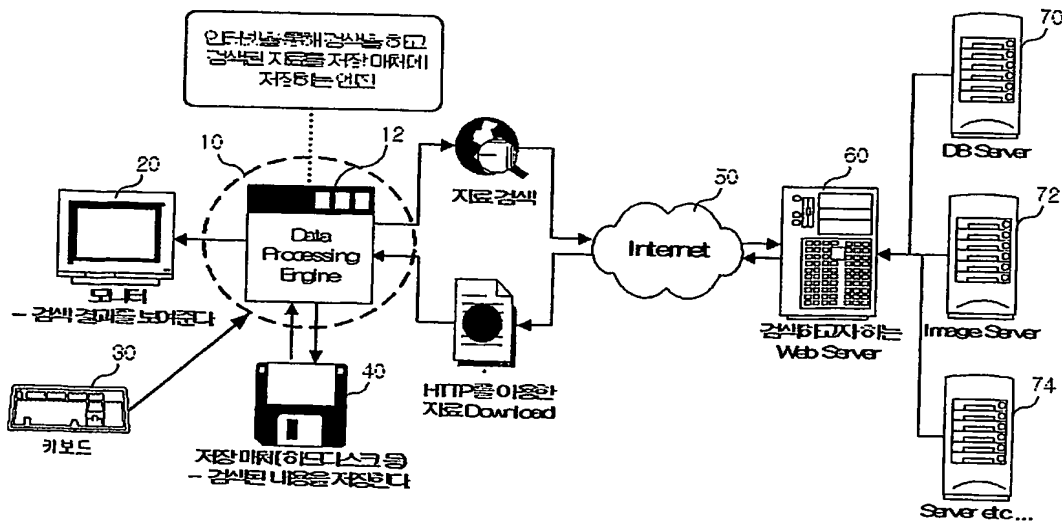
제1항에 있어서, 상기 로컬 저장매체는 플로피디스크, 하드디스크, 콤팩트디스크 또는 플래쉬 메모리 중에서 어느 하나인 것을 특징으로 하는 인터넷을 이용한 정보 자료 스크랩 방법.

【도면】

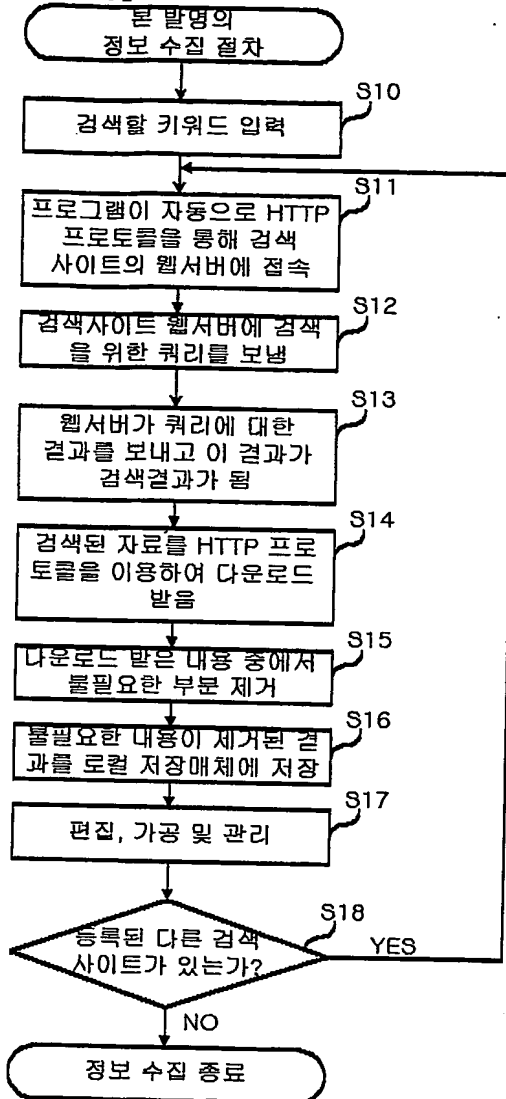
【도 1】



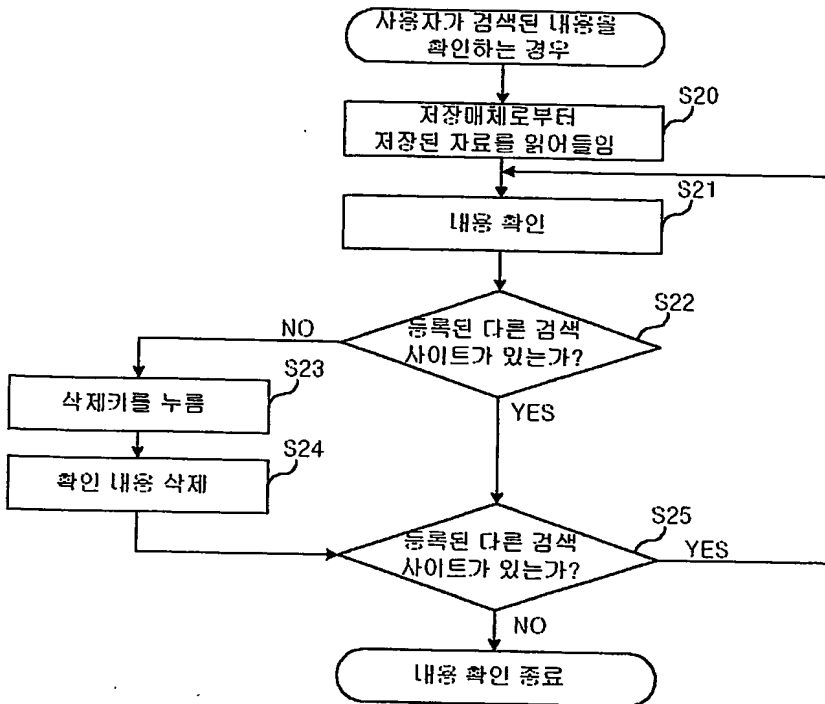
【도 2】



【도 3】



【도 4】



【도 5】

파일(F) 편집(E) 보기(V) 검색(S) 도구(T) 도움말(H)

검색 키워드 x

- 검색 키워드
- 썬니YNK
- 엔씨소프트
- 송해교
- 바둑기보
- 서유기

제 목	신 문 사	날 짜
[바둑] 한국기원, 1000만 바둑 전 통합DB ...	조선일보	2002-10-1
[기원] 한국기원, 1000만 바둑 전 통합DB ...	조선일보	2002-09-15
[바둑] 도요다현소배 아킬라, 마마로는 첫...	조선일보	2002-09-01
세계바둑 頂上에 오른 10대 소년李世鵬	조선일보	2002-09-01

기사 제목 :
관련 정보 : 조선일보에서 검색된 내용입니다.

●합솟와 대생

<제1보>(1~14)=이창호와 장주주(江鑄久)의 조합은 국제 대회 사상 처음이다. 항상 화제의 중심권에 놓여 온 두 기사는 오늘 어떤 합주(合奏)를 들려줄까. 결과를 미리 말하자면 기대했던 이상의 훌륭한 기보가 탄생했다. 승부의 속성은 기본적으로 대립이지

① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ ⑪ ⑫ ⑬ ⑭

Information

- * 키워드 "바둑기보"의 총 검색 시간 : 00시간 00분 32초
- 키워드 "바둑기보"의 검색이 완료되었습니다. - 2002년 10월 23일 (수), 22시 08분 51초
- 키워드 "바둑기보"의 검색을 시작합니다. - 2002년 10월 23일 (수), 22시 08분 19초
- 키워드 "서유기"의 총 검색 시간 : 00시간 00분 35초
- 키워드 "서유기"의 검색이 완료되었습니다. - 2002년 10월 23일 (수), 22시 05분 31초

Messages / Errors /

완료

【도 6】

키워드 등록

검색 키워드

키워드 등록

검색 키워드

새니YNK

엔씨소프트

중학교

바둑기보

서유기

검색하고자 하는 키워드를 입력하세요.

장엄마미템

<< 간단하...

검색 대상

☒ 제목 or 본문 ☐ 제목만 ☐ 본문만 ☐ 제목 and 본문

검색 기간 지정

검색 시작일 : ☒ 가능한 오래전 부터 ☐ 오늘로부터 1년 전 부터

날짜 지정 :

검색 종료일 : ☒ 오늘까지 ☐ 날짜 지정 :

확인 취소

키워드 "서유기"의 검색이 완료되었습니다. - 2002년 10월 23일 (수), 22시 08분 51초

Messages / Errors

Ready

【도 7】

키워드

검색 키워드

새니YNK

엔씨소프트

중학교

바둑기보

서유기

자료 검색 중...

현재 검색이 진행 중입니다...

전체 검색 상황

현재 사이트 검색 상황

현재 사이트 저장 상황

현재 사이트

검색된 자료 수

검색 취소

키워드 "서유기"의 검색을 시작합니다. - 2002년 10월 23일 (수), 22시 22분 35초

키워드 "바둑기보"의 총 검색 시간 : 00시간 00분 32초

키워드 "바둑기보"의 검색이 완료되었습니다. - 2002년 10월 23일 (수), 22시 08분 51초

키워드 "바둑기보"의 검색을 시작합니다. - 2002년 10월 23일 (수), 22시 08분 19초

키워드 "서유기"의 총 검색 시간 : 00시간 00분 35초

Messages / Errors

완료

【도 8】

검색 키워드 x

☒ 검색 키워드
☒ 씨리YNK
☒ 엔씨소프트
☒ 숭례교
☒ 바둑기보
☒ 서유기

【도 9】

제 목	신 문 사	날 짜
[바둑] 한국기원, 1080만 바둑 전 통합DB ...	조선일보	2002-10-1
[바둑] 한국기원, 1080만 바둑 전 통합DB ...	조선일보	2002-09-16
[바둑] 도요다-전소바 이길라, 아마로는 첫...	조선일보	2002-09-04
세계바둑 頂上에 오른 10代 소년 李世鵬	조선일보	2002-09-01

기사 제목:

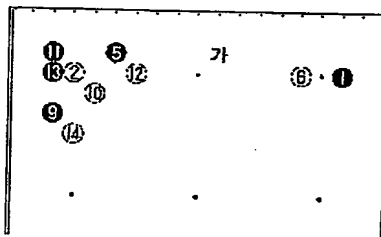
【도 10】

기사 제목:

관련 정보: 조선일보에서 검색된 내용입니다.

● 합솟와 대생

<제1보>(1~14)=이창호와 장주주(江鑄久)의 조합은 국제 대회 사상 처음이다. 항상 화제의 중심권에 놓여 온 두 기사는 오늘 어떤 합주(合奏)를 들려줄까. 결과를 미리 말하자면 기대했던 이상의 훌륭한 기보가 탄생했다. 승부의 속성은 기본적으로 대립이지



【도 11】

바둑기보.doc - Microsoft Word

파일(F) 편집(E) 보기(V) 삽입(I) 서식(O) 도구(T) 표(A) 창(W) 도움말(H)

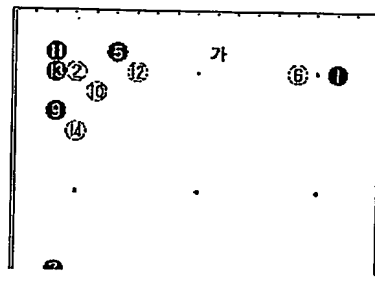
100% 표준+굵게 12 가

카드 서비스 중 가장 눈길을 끄는 것은 카메라를 통한 영상 변환 모듈 방식으로 전 기전을 중계한다는 점. 기록자의 입력을 생략하는 디지털 방식을 채택, 시간과 정확도를 훨씬 높였다. 지난 4월 특허를 따낸 기보(棋譜) 생성 장치 시스템에 의한 개인 기보의 자동 DB화도 획기적이다. 이밖에 한국기원 주관 대회 참가비와 '월간 바둑' 구독료, 아마 단 인허 수수료 등의 할인 특전도 부여할 예정. 문의 (02)730-3774.

【기왕전】 본선 2회전 제 6국 ○장주주●이창호(1보)
(조선일보)

●합죽와 대생

(1~14)=이창호와 장주주(江鑄久)의 조합은 국제 대회 사상 처음이다. 항상 화제의 중심권에 놓여 온 두 기사는 오늘 어떤 합죽(合奏)을 들려줄까. 결과를 미리 말하자면 기대했던 이상의 훌륭한 기보가 탄생했다. 승부의 속성은 기본적으로 대립이지만, 파트너에 따라선 멋진 화음(和音)이 펼쳐지기도 한다.



1 페이지 1 구역

한국어

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.